

This is the **accepted version** of the book part:

Fernandez, F. V.; Roca, E.; Saraza, P.; [et al.]. «Strategies for parameter extraction of the time constant distribution of time-dependent variability models for nanometer-scale devices». A: 2023 19th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD). 2023. DOI 10.1109/SMACD58065.2023.10192206

This version is available at <https://ddd.uab.cat/record/291513>

under the terms of the  **CC BY-NC-ND** license

Strategies for parameter extraction of the time constant distribution of time-dependent variability models for nanometer-scale devices

F.V. Fernandez¹, E. Roca¹, P. Sarazá², J. Martin-Martinez³, R. Rodriguez³, M. Nafria³ and R. Castro-Lopez¹

¹ Instituto de Microelectrónica de Sevilla, IMSE-CNM (CSIC and Universidad de Sevilla), Sevilla, Spain

² IMEC, Leuven, Belgium

³ Electronic Engineering Department (REDEC) group, Universitat Autònoma de Barcelona (UAB) Barcelona, Spain

Abstract—Time-dependent variability phenomena are stochastic and discrete for nanometer-scale technologies, and, hence, must be statistically characterized. These phenomena are attributed to the emission and capture of charges in device defects. This paper explores two different strategies to extract, from experimental data, the distribution parameters of the time constants of the defects. It delves into the accuracy of each strategy, showing how the extraction strategy can have a huge impact on the accuracy and the amount of characterization data required, and, therefore, on the amount of (expensive) characterization time in the lab.

Keywords— time-dependent variability, modeling, parameter extraction, characterization

I. INTRODUCTION

Time-dependent variability (TDV) phenomena in nanometric CMOS technologies, such as Bias Temperature Instability (BTI), Hot Carrier Injection (HCI), or Random Telegraph Noise (RTN), has become a subject of increasing concern. BTI and HCI result in a gradual shift of transistor parameters over time, e.g., an increase in the absolute value of the threshold voltage, V_{th} [1]. These variations happen over months and years under nominal operation conditions and eventually lead to device failure. Aging (due to HCI and BTI), considered to have a permanent and a (statistically distributed, for small devices) recoverable component, has thus become a major concern for long-term circuit functionality. The origin of these phenomena is attributed to the capture/emission of charge carriers in/from defects in the transistors, which, in nanometer-scale transistors, induce discrete shifts of the threshold voltage and, hence, current shifts [2]. The parameters associated with these defects, namely the number of defects per transistor, their emission time τ_e and capture time τ_c , as well as the magnitude of the induced threshold voltage shifts, change from device to device, their values following probability distributions. Defect-centric models, such as the Probabilistic Defect Occupancy (PDO) model [3], have been proposed to account for the stochasticity of TDV.

Since it is not feasible to characterize transistors over years, the typical aging characterization procedure uses accelerated aging tests, in which temperature and/or the drain voltage and/or the gate voltage are raised above their nominal values for a much shorter time, i.e., the stress time. For instance, elevated voltages are applied to the devices during several periods of stress, followed by current measurements at low voltages to evaluate the impact of the stress on the device characteristics (known as MSM procedures). Then, the results obtained under accelerated stress conditions can be extrapolated to normal operating conditions.

This work was supported by grant ProyExcel_00536 funded by Consejería de Universidad, Investigación e Innovación de Junta de Andalucía. It was also supported by grants PID2019-103869RB-C31/C32 funded by MCIN/AEI/10.13039/501100011033 and by grant TED2021-131240B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the “European Union NextGenerationEU/PRTR”.

Due to the stochastic nature of these TDV phenomena, a large number of devices must be characterized. To perform a fast statistical characterization of the TDV effects, array structures with thousands of transistors are the best option, as they enable parallelizing the aging tests by stressing several devices at the same time. Accurate statistical characterization of RTN, BTI and HCI in a single array-based chip was reported in [4]. This 65-nm CMOS chip as well as the automated measurement setup that was built around it [5] will be exploited in this paper. A methodology to extract the number of defects, the time-to-emissions and the time-to-captures and current shifts of each device from its experimental current trace was developed in [6].

The existing literature has mainly focused on the analytical formulation of the probability distribution functions of the device parameters and the characterization strategies, but limited attention has been paid to how exactly to attain the distribution parameters from the laboratory characterization results. This paper explores different strategies to extract, from experimental data, the parameters of the stochastic distributions followed by the time constants of the defects. The paper is structured as follows. Section II introduces the mathematical modeling of the defect distribution and the experiment setup. Two strategies for parameter extraction are described in Sections III and IV. An experimental comparison is performed in Section V.

II. MATHEMATICAL FORMULATION

As stated above, the usual aging characterization procedure resorts to consecutive periods of stress (where the device characteristics are degraded) and recovery periods (where the measurement of the device current is performed during a certain time window and the degraded device partially recovers from its degradation). The recovery of the transistor current, in the form of discrete increments, corresponds to the emission of charges from defects that were previously captured during the stress phase. As it is illustrated in Fig. 1, the number of emission events, their amplitude and the time instant when these events take place, i.e., the time-to-emission t_e , can be all obtained from the current trace. The time-to-emission data of each defect, which are but samples of its emission time constant, will be of relevance for the analysis reported here.

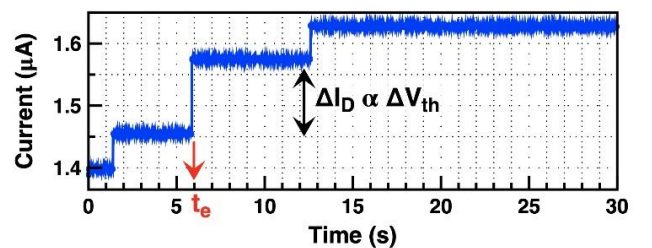


Fig. 1. Current trace in a recovery phase with three emission events.

For a charge emission event to be observed during the recovery phase, three conditions must occur: (1) the defect must exist, (2) it must be occupied at the beginning of the measurement phase and (3) it must be emitted within the experimental measurement window. The probability that a defect, characterized by a given pair of time constants (τ_e , τ_c), exists is determined by the probability density function P_{def} . Though a bivariate lognormal distribution will be considered here [7]:

$$P_{def}(\tau_e, \tau_c) = \frac{1}{2\pi\tau_e\tau_c\sigma_{\tau_e}\sigma_{\tau_c}\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(\ln\tau_e - \mu_{\tau_e})^2}{\sigma_{\tau_e}^2} + \frac{(\ln\tau_c - \mu_{\tau_c})^2}{\sigma_{\tau_c}^2} - \frac{2\rho(\ln\tau_e - \mu_{\tau_e})(\ln\tau_c - \mu_{\tau_c})}{\sigma_{\tau_e}\sigma_{\tau_c}}\right]} \quad (1)$$

the results in this paper are equally applicable to any other distribution. Parameters μ_{τ_e} and μ_{τ_c} in (1) are the mean values of the emission time's logarithm $\ln\tau_e = \log(\tau_e)$ and capture times' logarithm $\ln\tau_c = \log(\tau_c)$, σ_{τ_e} and σ_{τ_c} are the standard deviations and ρ is the correlation coefficient. An example of defect distribution function is plotted in Fig. 2a.

The probability that a defect is occupied at any time instant can be established as the probability that the defect was empty in a previous time instant and captures a charge in the time lapse between both time instants plus the probability that it was occupied and does not emit a charge in the same time lapse. In differential form, this can be formulated as:

$$\frac{dP_{occ}(t)}{dt} + P_{occ}(t)\left(\frac{1}{\tau_e} + \frac{1}{\tau_c}\right) - \frac{1}{\tau_c} = 0 \quad (2)$$

Notice that the time constants do change with the gate and drain voltages. In this work, an exponential dependence will be considered:

$$\begin{aligned} \tau_e &= \tau_{e0} 10^{\beta_e(|V_{gs}| - |V_{gsref}|)} 10^{\gamma_e(|V_{ds}| - |V_{dsref}|)} \\ \tau_c &= \tau_{c0} 10^{\beta_c(|V_{gs}| - |V_{gsref}|)} 10^{\gamma_c(|V_{ds}| - |V_{dsref}|)} \end{aligned} \quad (3)$$

Parameters τ_{e0} and τ_{c0} correspond to the time constants for $|V_{gs}| = |V_{gsref}|$ and $|V_{ds}| = |V_{dsref}|$. Here, we consider the

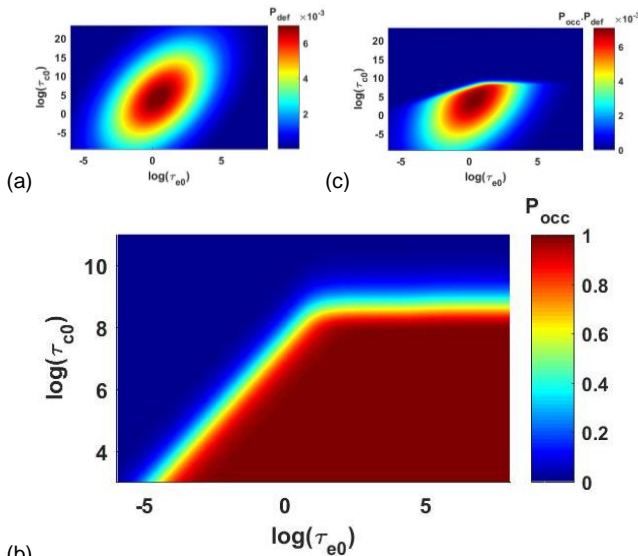


Fig. 2. (a) Defect distribution; (b) Probability of occupation after 10,000s of stress at $|V_{gs}| = 2.5V$; (c) Product of P_{def} and P_{occ} .

reference voltages as those used during the recovery phase. As the biasing conditions remain constant during each stress/recovery phase, (2) can be easily solved analytically to get the probability of occupation at the end of each phase ($t = t_f$) given the probability at the beginning of that phase $t = 0$:

$$P_{occ}(t_f | \tau_e, \tau_c) = \frac{\tau_e}{\tau_e + \tau_c} + \left(P_{occ}(t = 0 | \tau_e, \tau_c) - \frac{\tau_e}{\tau_e + \tau_c} \right) e^{-\left(\frac{1}{\tau_e} + \frac{1}{\tau_c}\right)t_f} \quad (4)$$

As an example, Fig. 2b plots the probability of occupation at the end of a stress phase of 10,000s at $|V_{gs}| = 2.5V$.

The product of (1) and (4) at the end of the stress period, plotted in Fig. 2c, provides the charged defects that can experience an emission event. However, this does not imply that an emission will be observed during the measurement. The emission of charges is commonly modeled by a Markov process, whose probability density function (pdf) is given by (with direct and logarithm variables):

$$P(t_e) = \frac{1}{\tau_e} e^{-\frac{t_e}{\tau_e}} \quad P(\ln t_e) = \frac{t_e \ln(10)}{\tau_e} e^{-\frac{t_e}{\tau_e}} \quad (5)$$

where $\ln t_e = \log(t_e)$. Fig. 3 shows the representation of the pdf for a defect with emission time constant $\tau_e = 100s$. It can be observed that the time at which an emission event occurs (the time-to-emission t_e) in a defect with a given time constant τ_e spans over several decades.

The possible causes for not observing or considering an emission event of an occupied defect are:

- The charge is emitted before the first measurement.
- The charge is emitted after the end of the measurement window
- The charge is emitted during the measurement window, but it is captured again before the end of the measurement window. This capture event may happen in a shorter time frame than the temporal resolution of the measurement equipment; hence, the emission event is simply not detected. But it may also happen along the measurement window. Although the emission event is detected in this case, there will be at least one emission and capture event, and probably many more, and it is typically identified as RTN. Although the emission event is detected, joint mathematical treatment with the other emission events is very complex and they are usually discarded in this modeling process.

The probability of occurrence of an emission event without a subsequent capture at a time t_e of the measurement window is given by the probability that a charge is emitted

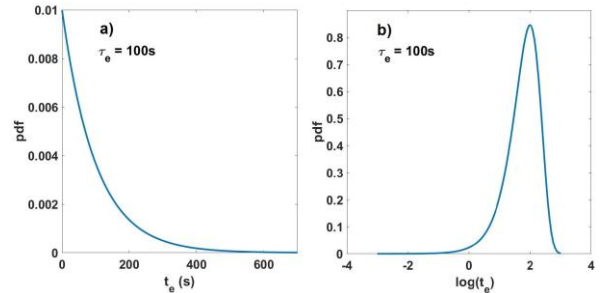


Fig. 3. Pdf of (a) time-to-emissions and (b) logarithm of time-to-emissions of a defect with emission time constant $\tau_e = 100s$.

after the beginning of the measurement window multiplied by the probability that such defect is not captured again before the end of the measurement window [7]:

$$P_{see}(t_e|\tau_e, \tau_c) = \begin{cases} 0 & \text{if } t_e < t_{min} \\ \frac{1}{\tau_e} e^{-\frac{t_e}{\tau_e}} \cdot e^{-\frac{t_{max}-t_e}{\tau_c}} & \text{if } t_{min} < t_e < t_{max} \\ 0 & \text{if } t_e > t_{max} \end{cases} \quad (6)$$

III. STRATEGY 1: MATCHING THE EMISSION TIME CONSTANT DISTRIBUTION

The first strategy consists in determining the distribution parameters that better fit the emission time constants of defects occupied after a certain stress time. Equation (6) provides the distribution of time-to-emissions of a defect characterized by time constants τ_e and τ_c . Integrating (6) in the measurement window $[t_{min}, t_{max}]$ provides the probability that an emission event of a defect is observed during such window:

$$P_{vis}(\tau_e, \tau_c) = \int_{t_{min}}^{t_{max}} P_{see}(t_e|\tau_e, \tau_c) dt_e \\ = \frac{\tau_c}{\tau_e - \tau_c} \left[e^{-\frac{t_{max}}{\tau_e}} - e^{-\frac{t_{min}}{\tau_e}} - e^{-\frac{t_{min}}{\tau_e}} e^{-\frac{t_{max}-t_{min}}{\tau_c}} \right] \quad (7)$$

Fig. 4a illustrates the representation of $P_{vis}(\tau_e, \tau_c)$ for the experimental conditions in this work: the first sample at the measurement window is taken after 2ms ($t_{min} = 2\text{ms}$) and the measurement window lasts 100s ($t_{max} = 100\text{s}$). The product of (7), plotted in Fig. 4a, and (4), plotted in Fig. 2b, yields the probability that an existing defect is occupied at the end of the stress period and an emission event is detected during the measurement window. This probability is illustrated in Fig. 4b. Then, since the MSM measurement procedure does not monitor capture events, but only emissions, it is possible to integrate the product of (1), (4) and (7) along the τ_c axis to obtain the probability that a defect with emission time τ_e exists, is occupied at the beginning of the measurement window, and experiences an emission event during that window:

$$P_{lab-vis-\tau_e}(\tau_e) = \int_0^{+\infty} P_{def}(\tau_e, \tau_c) P_{occ}(t_f|\tau_e, \tau_c) P_{vis}(\tau_e, \tau_c) d\tau_c \quad (8)$$

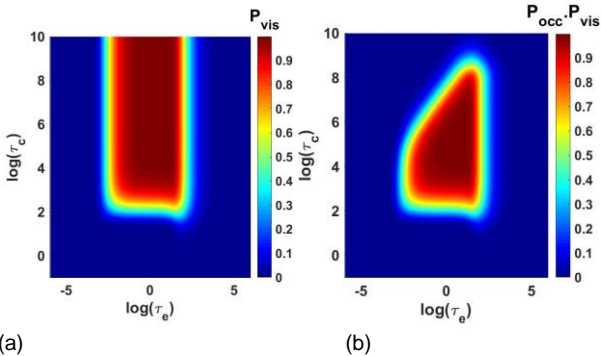


Fig. 4. (a) Probability of visualization in the measurement window; (b) Probability of visualization of defects occupied at the end of the stress period.

The goal is to fit this distribution, plotted in Fig. 5, to the experimental results. However, the experimental current traces do not provide samples of the distribution of the emission time constant but a set of time-to-emissions of all defects. Therefore, it is necessary to characterize the emission time constants from the set of time-to-emissions of each defect [8]. This implies that the MSM procedure must be repeated a number of times N_{SM} so that the emission of a charge from a given defect can be detected several times, each with a different time-to-emission. If an emission were observed each of the N_{SM} times an MSM procedure is performed, the emission time constant could be obtained by simply averaging the emission times. But one only has to look at Fig. 3 to realize that many emissions will occur outside the experimental window. Therefore, some more sophisticated methods are required. Here, we will consider an approach that uses the logarithm of the time-to-emissions lt_e and fits the cumulative density function (cdf) obtained from the integration of the right equation in (5). Fig. 6a shows the fitting of the cdf built with the 100 emission events of a defect with a given time constant. However, there can be very fast emission events that occur before the measurement window starts and this could lead to an incorrect calculation of the occupancy probability if those undetected events are not accounted for. Nevertheless, this can be easily considered because the emission event of a given defect is observed less than N_{SM} times. The absence of the emission event may happen because the emission occurs before the start of the measurement window but also because it occurs after the end of the measurement window. Identifying which is the case is essential. This can be easily accomplished by assessing if the observed emission events are closer to the start (Fig. 6b) or the end of the measurement window (Fig. 6c). However, in both cases, accuracy will decrease since fewer points are available for the fitting. Fig. 3b shows that this happens for around three decades around the start and end of the measurement window.

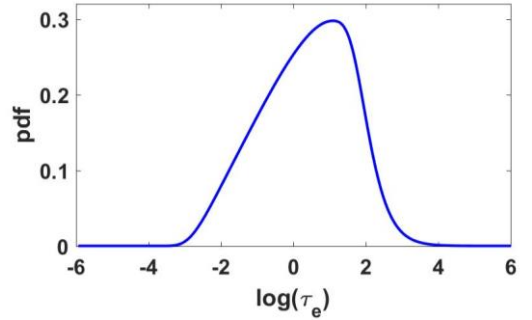


Fig. 5. Probability density function of emission time constants.

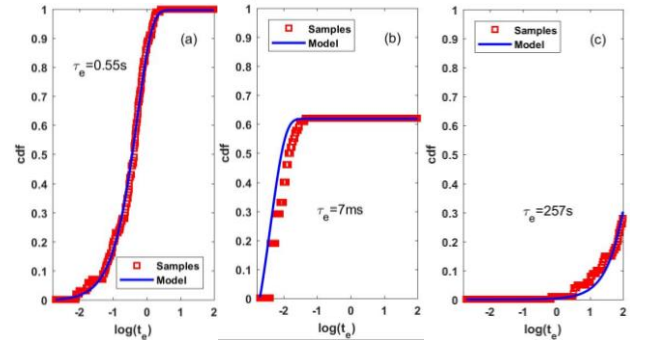


Fig. 6. Fitting the cdf of time-to-emissions of a single defect to get the emission time constant.

IV. STRATEGY 2: MATCHING THE TIME-TO-EMISSION DISTRIBUTION

This strategy consists in directly matching the cdf of time-to-emissions observed in the laboratory and does not need the determination of the emission time constant of each defect from a number of time-to-emissions in the same device [7]. The basic idea is that the emission events that are detected during the measurement window are given by the product of (1), (4) and (6):

$$P_{labsee}(t_e, \tau_e, \tau_c) = P_{def}(\tau_e, \tau_c) \cdot P_{occ}(t = t_0) \cdot P_{see}(t_e | \tau_e, \tau_c) \quad (9)$$

Then, it is possible to obtain the probability density function of the time-to-emissions t_e by integrating the three-dimensional distribution in (10) for every possible value of τ_e and τ_c :

$$P_{lab_vis_te}(t_e) = \iint P_{labsee}(t_e, \tau_e, \tau_c) d\tau_e d\tau_c \quad (10)$$

The corresponding cdf can be easily obtained from the integration of equation (11), and this cdf can be directly fitted to the cdf of the experimental time-to-emissions.

V. EXPERIMENTAL COMPARISON

The ultimate goal of the parameter extraction process is to get the parameter values in (1) and (3) that better match the experimental measurements. This can be formulated as the minimization of the differences between model evaluation and experimental data. Such kind of minimization problem is usually solved by applying stochastic optimization algorithms. As the main driver for any optimization algorithm is the difference between the model evaluation and the experimental data, we will consider the accuracy in such difference evaluation as the comparison criterion between the different strategies.

Fig. 7 performs the comparison for the approaches explored in this paper. In all cases, 400 devices with known distribution functions were used and, hence, the differences between model evaluation and experiment should be those intrinsic to the finite set of experimental data. In Fig. 7a, 100 MSM cycles were performed to enable the calculation of time constants, whereas in Fig. 7b a single MSM cycle was required. This is already a huge difference since the experimental data in Fig. 7b would require roughly 11 hours of lab measurements whereas Fig. 7a requires 46 days. But it can be observed that not only Strategy 2 is faster and requires fewer data, but also much more accurate. The reasons for the important deviations in Fig. 7a must be sought in the implicit assumption in Strategy 1: the defects are always either occupied or empty at the start of any recovery phase of the 100 MSM cycles [8]. It can be observed in Fig. 2b that there is a wide transition region between occupied ($P_{occ} = 1$) and empty defects ($P_{occ} = 0$), where there is a certain probability that the defect is charged at the end of the stress period. In case it is empty, an emission event during the recovery phase will obviously not happen. However, the approach in Strategy 1 cannot distinguish if the defect was not initially charged or the emission occurred outside the measurement window, and incorrectly considers that there are more defects close to the limits of the measurement window. Therefore, Strategy 1 can

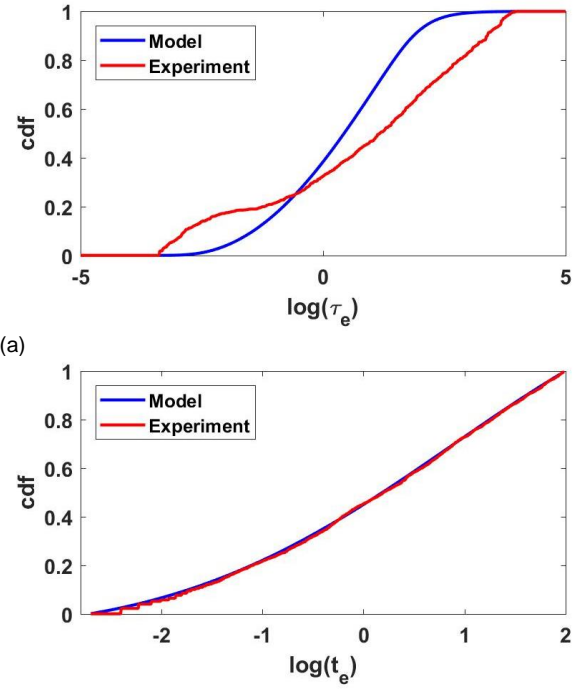


Fig. 7. (a) Comparison of cdf of emission time constants with Strategy 1, and (b) comparison of cdf of time-to-emissions with Strategy 2.

lead to large errors in the parameter extraction of the TDV models.

VI. CONCLUSIONS

This paper studies the impact of two different parameter extraction strategies on the accuracy of the fitting results and the amount of characterization data, which is directly related to the characterization cost. For the first time, it is identified that the assumption that the defects are always occupied or empty at the end of the stress period is behind the differences in accuracy of both techniques. Strategy 2 is shown to be insensitive to such assumption.

REFERENCES

- [1] T. Grasser et al, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," In IEEE International Reliability Physics Symposium (IRPS), pp. 16-25, 2010.
- [2] T. Grasser et al, "A unified perspective of RTN and BTI," In IEEE International Reliability Physics Symposium (IRPS), pp. 4A-5, 2014.
- [3] J. Martin-Martinez et al. "Probabilistic defect occupancy model for NBTI". In IEEE Inter. Reliability Physics Symposium, XT-4, 2011.
- [4] J. Diaz-Fortuny et al, "A versatile CMOS transistor array IC for the statistical characterization of time-zero variability, RTN, BTI and HCI," IEEE J. Solid-State Circuits, vol. 54, no. 2, pp. 476-488, 2019.
- [5] J. Diaz-Fortuny et al, "Flexible setup for the measurement of CMOS time-dependent variability with array-based integrated circuits," IEEE Trans. Instrum. Meas., vol. 69, no. 3, pp. 853-864, Mar. 2020.
- [6] P. Saraza-Canflanca et al, "A robust and automated methodology for the analysis of time-dependent variability at transistor level," Integration, the VLSI Journal, vol. 72, pp. 13-20, 2020.
- [7] P. Saraza-Canflanca et al, "Determination of the time constant distribution of a defect-centric time-dependent variability model for sub-100-nm FETs," IEEE Trans. on Electron Devices, vol. 69, no. 10, pp. 5424-5429, Oct. 2022.
- [8] H. Reisinger et al, "The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress," In IEEE Int. Reliability Physics Symp. (IRPS), pp. 7-15, 2010.